Input Variable and NN cuts

Richard St. Denis University of Glasgow/CDF

26 November 2007

Correlations in NN Output and Input Variables

- Control Region Method
- Application of Control Region and improved statistical evaluations

Correlations in NN Output and Input Variables

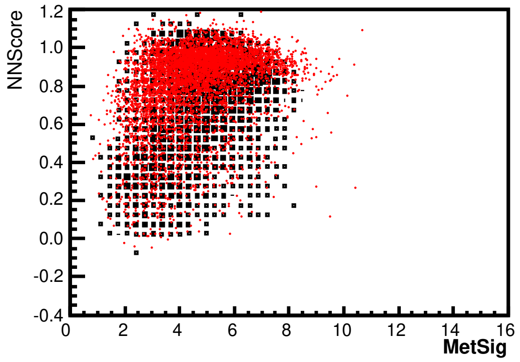
- Method
  - Look at NN vs Input for HWW and backgrounds.
  - Find point in NN score above which majority of HWW resides. This is the critical score.
  - Determine range of Input variable where majority of HWW resides. This defines the input variable sensitive region.
  - Determine if there are background events populating this sensitive region and for which the background is below the critical score. A band at low NN score is best, indicative of another cut that forms a control region.

- Examine the data input distributions, blanking out events populating the region above the critical score. Compare to the background.

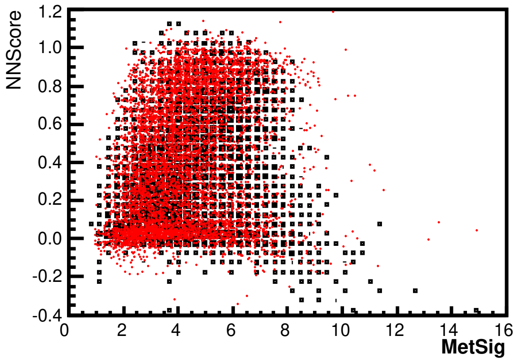
- Here we will look at MetSig and the six distributions with most separation of signal and background.
  - Look at NN vs Input variable for HWW, WW and Data for two examples.
  - Evaluate if there is sufficent background in the sensitive region.
  - Summarize results;reamining graphs for variables with most separation in backup.

High S/B, 0,1 Jet, NN vs MetSig

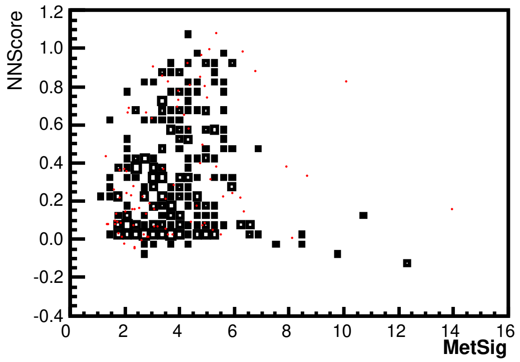
Higgs



WW

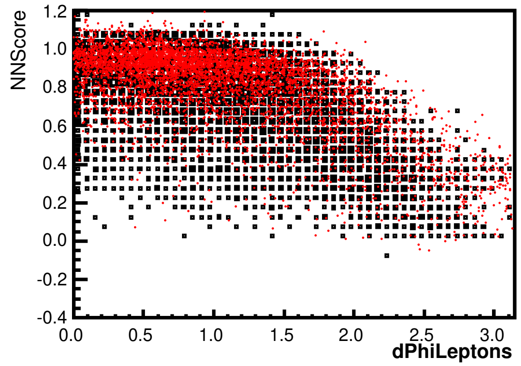


Data

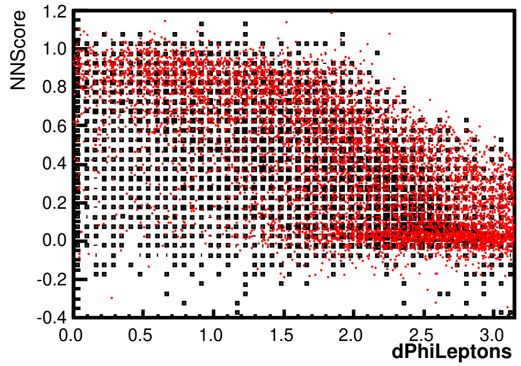


High S/B, 0,1 Jet, NN vs DPhiLeptons

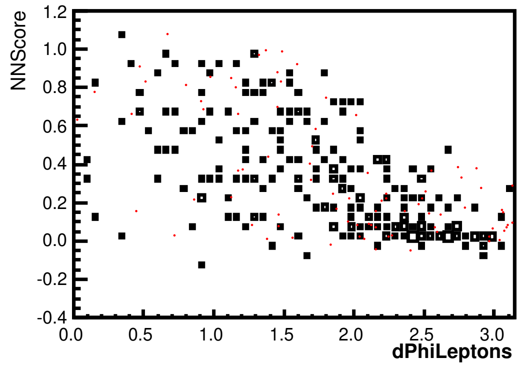
Higgs



WW



Data



Summary

- **MetSig** Reasonable coverage of signal region with an interesting band in WW at NN=0)at least in 1 jet) that covers signal region nicely.
- **LRHWW** Some coverage at highest NN scores: not very good and it should not be!
- **dRLeptons** Coverage for small DR, signal region, is not very good, especially for 1 jet. This is a very powerful discriminant and as such is hard to find a good low NN score region with low DR.
- **dPhiLeptons** Again, as in dRLeptons, coverage is not very good at low dPhi.
- **dimass** Good coverage in signal region.
- **MetSpec** Good coverage with a band at low NN score

(at least in 1 jet).

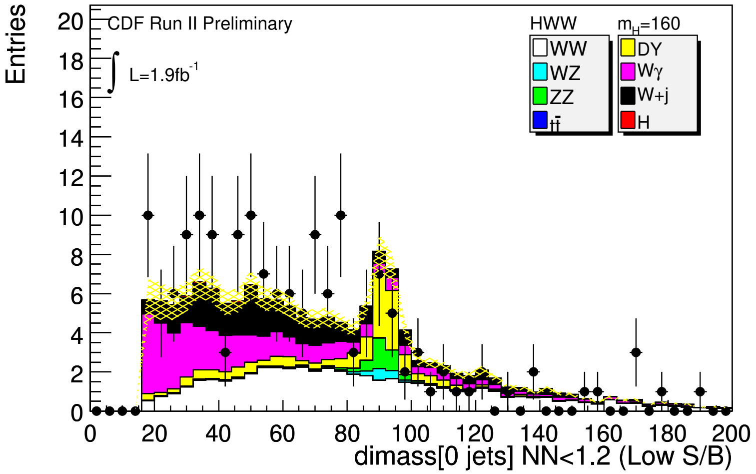
- **Met** Good coverage with a band at low NN score (at least in 1 jet).

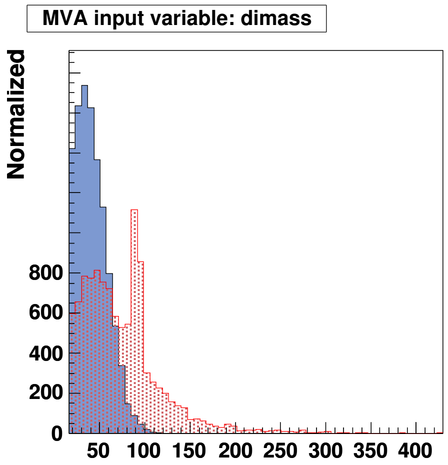
Application to Most Separated Variables

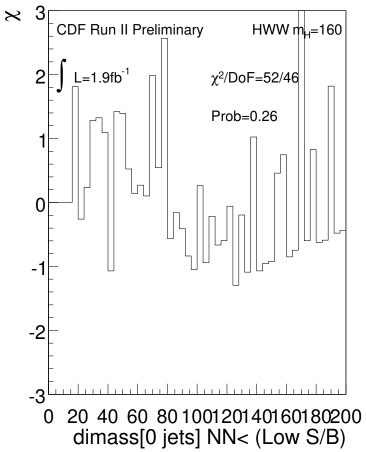
- Looked at distributions and χ^2 probabilities. Fixed them so that the DOF was counted for non-zero MC bins. Also, estimate error on data events from expected number of MC events.
- Picked out distributions that look bad in χ^2 and KS and are in the top six. Also looked at MetSig.
- Separated out distributions of by cutting on NNScore so that 80% of the signal is included or excluded when NN score is above/below cut. Re-evaluate KS Probability and χ^2.

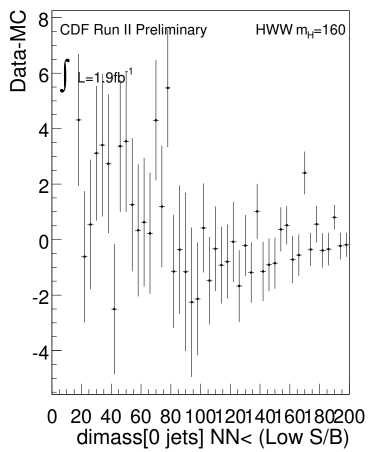
	High S/B 0 Jet			
Discriminant	KS Prob	χ^2/DOF	χ^2 Prob	Comment
LRHWW	0.05	62/50	0.11	Visual ok
dRLeptons	0.18	33/36	0.62	
dPhiLeptons	0.44	37/50	0.91	
dimass	0.58	63/46	0.05	Visual Ok
MetSpec	0.53	68/47	0.03	3 Evt in tail
Met	0.65	65/47	0.04	3 Evt in tail
	Low S/B 0 Jet			
Discriminant	KS Prob	χ^2/DOF	χ^2 Prob	Comment
LRHWW	0.04	65/50	0.08	χ^2 grow after 0.5
dRLeptons	0.23	33/42	0.84	
dPhiLeptons	0.76	56/50	0.25	
dimass	0.00	52/46	0.26	Show: High at Low Mass. See Signal
MetSpec	0.43	30/47	0.98	
Met	0.35	36/47	0.87	

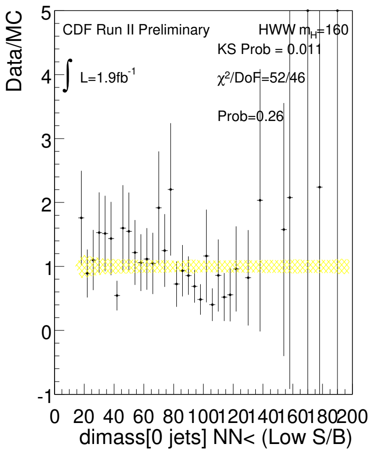
M_{ll}: Low S/B 0 Jet



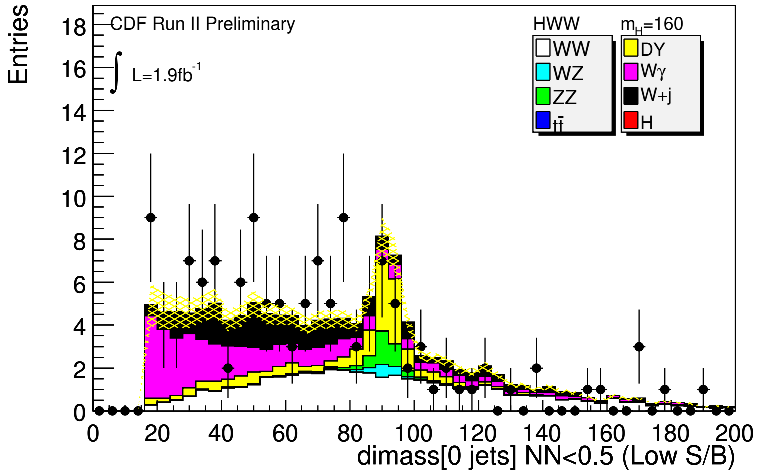


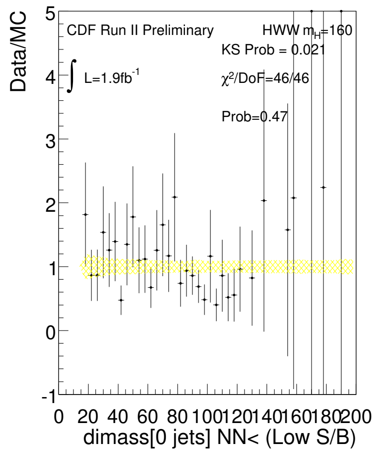


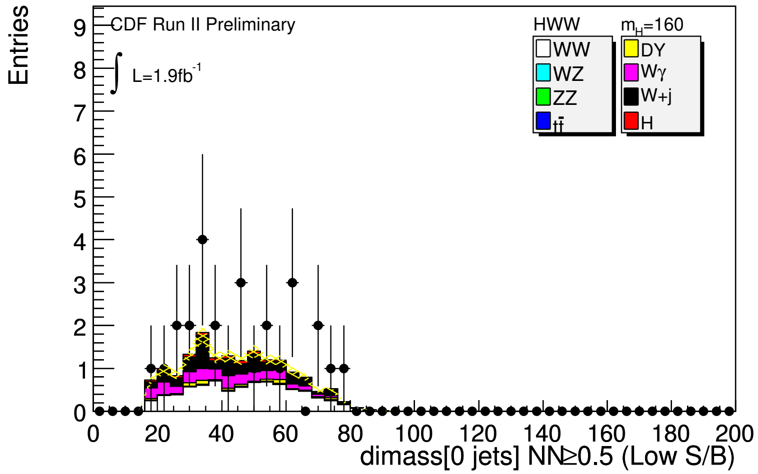


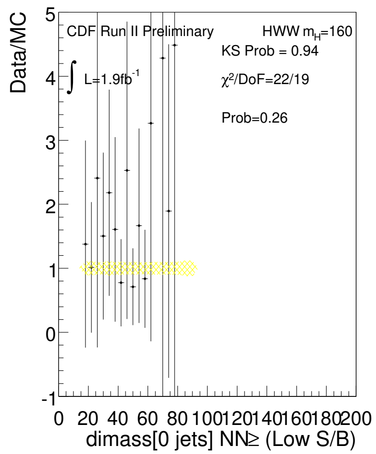


M_{ll}: Low S/B 0 Jet



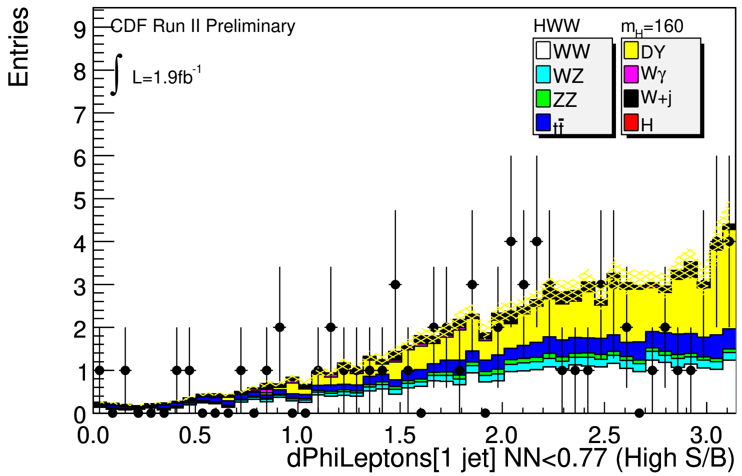


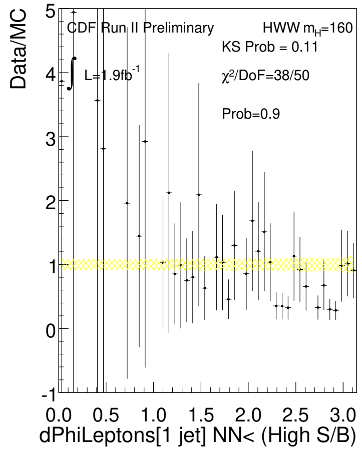


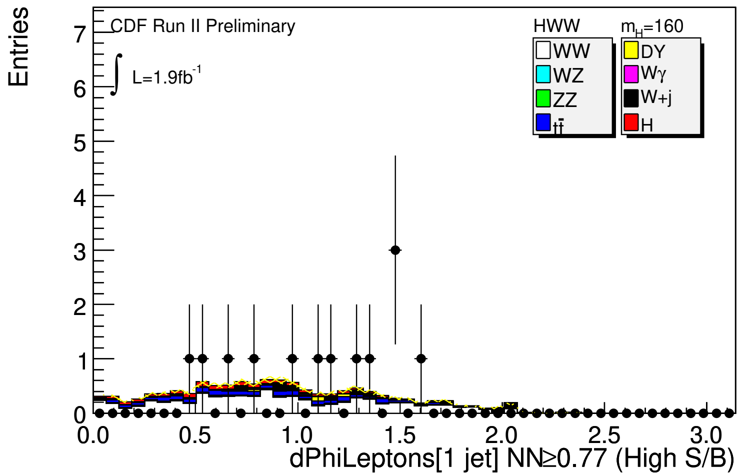


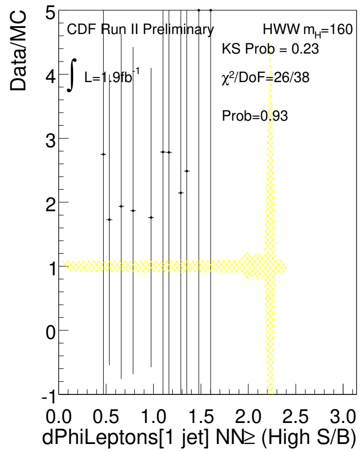
	High S/B 1 Jet			
Discriminant	KS Prob	χ^2/DOF	χ^2 Prob	Comment
LRHWW	0.12	44/50	0.73	
dRLeptons	0.07	39/36	0.35	ok:KS?
dPhiLeptons	0.07	38/50	0.88	Show: Data Flatter high ϕ
dimass	0.09	39/46	0.76	Show Sig
MetSpec	0.79	28/47	0.99	
Met	0.91	26/47	0.99	
	Low S/B 1 Jet			
Discriminant	KS Prob	χ^2/DOF	χ^2 Prob	Comment
LRHWW	0.04	59/50	0.18	
dRLeptons	0.28	48/43	0.27	
dPhiLeptons	0.01	77/50	0.01	Show: Data Flatter
dimass	0.41	50/46	0.32	
MetSpec	0.14	33/46	0.92	
Met	0.11	24/47	1	

dPhiLeptons: High S/B 1 Jet

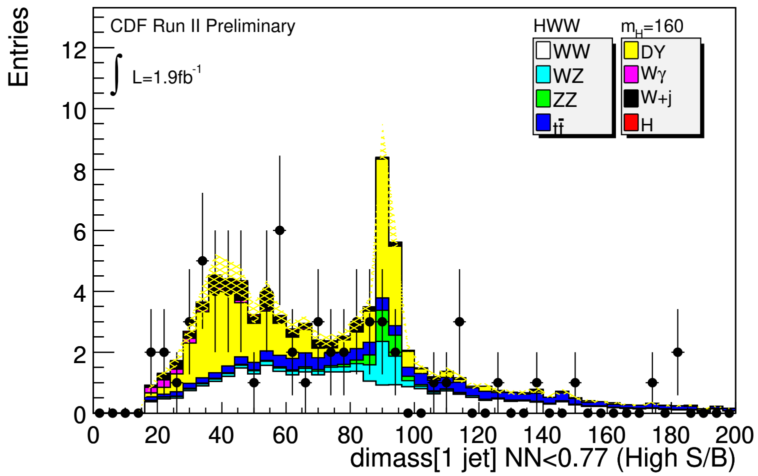


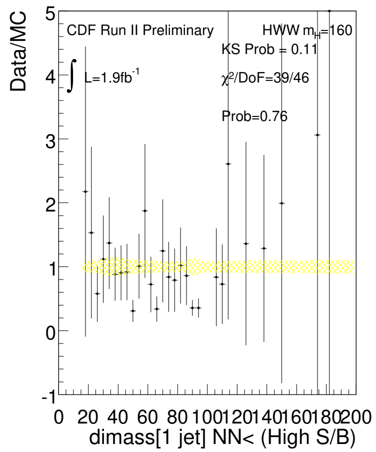


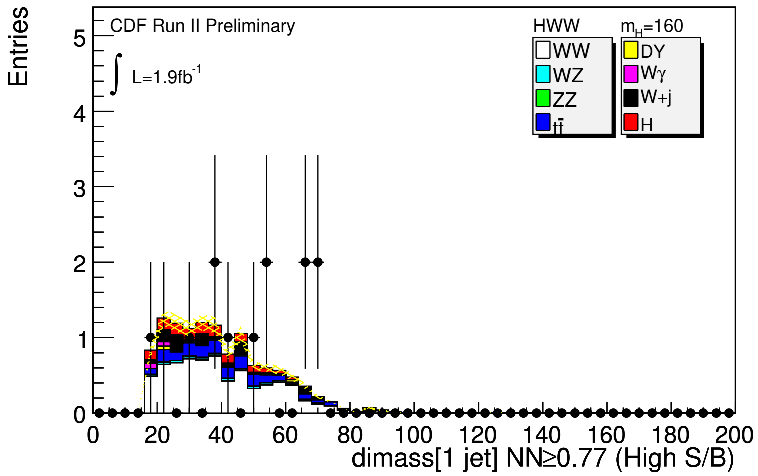


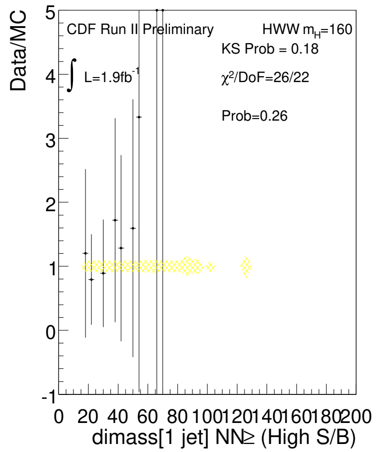


dimass: High S/B 1 Jet

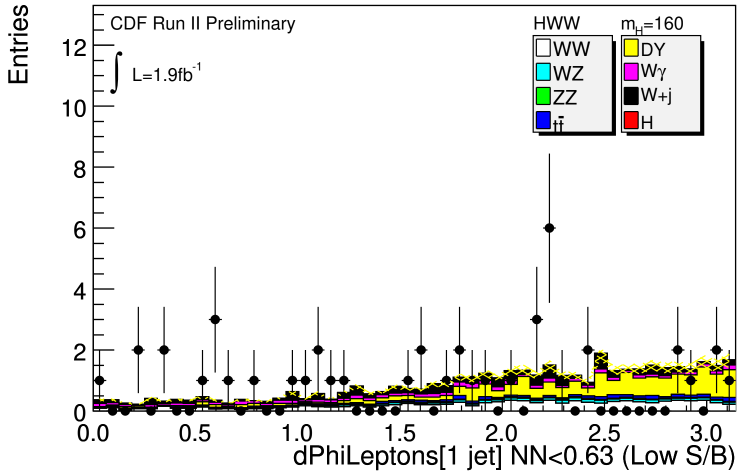


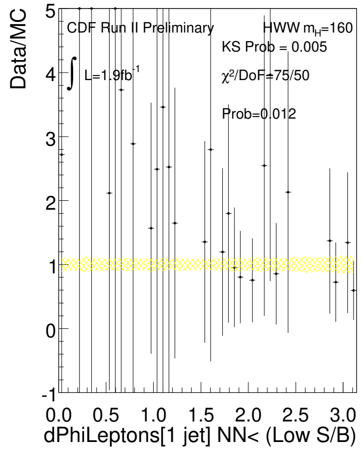


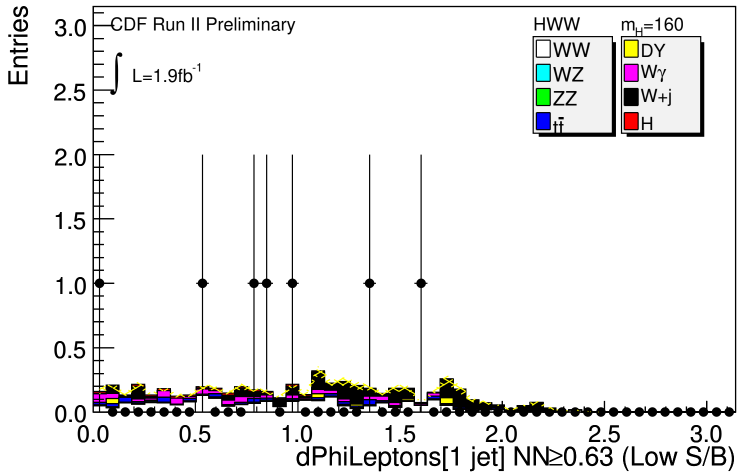


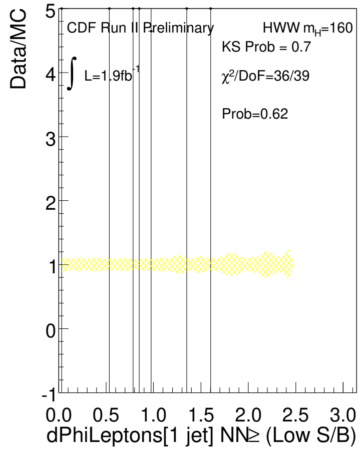


dPhiLeptons: Low S/B 1 Jet

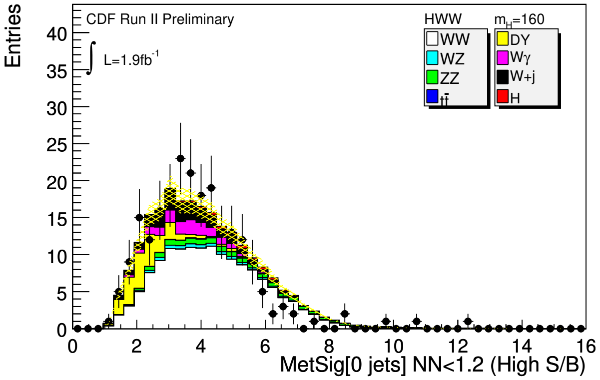


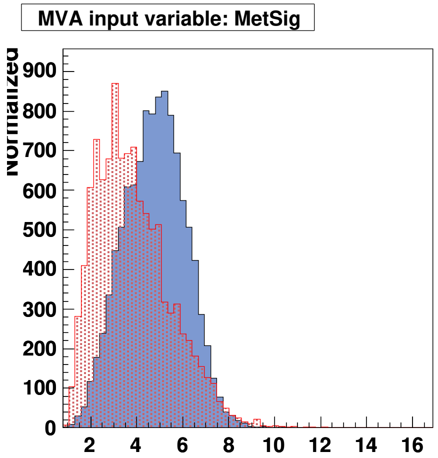


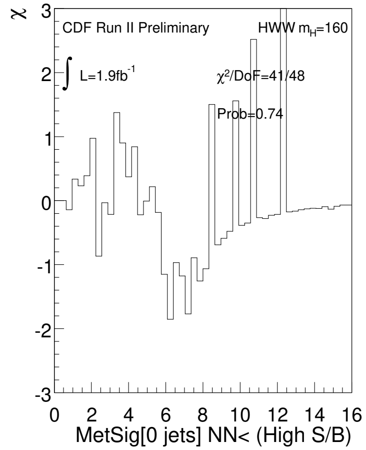


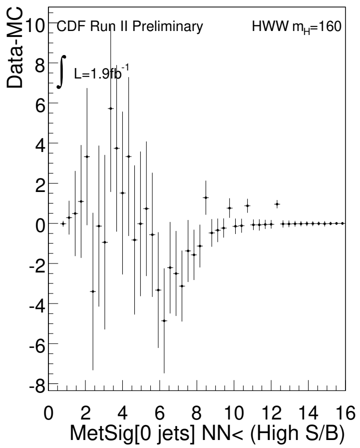


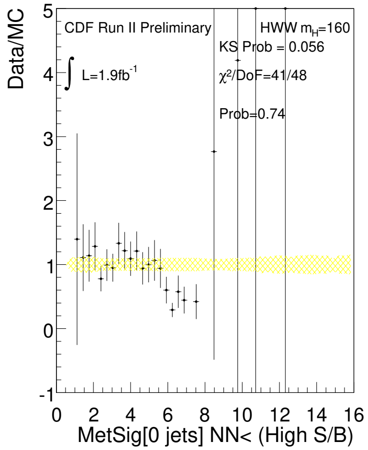
MetSig Sig Region



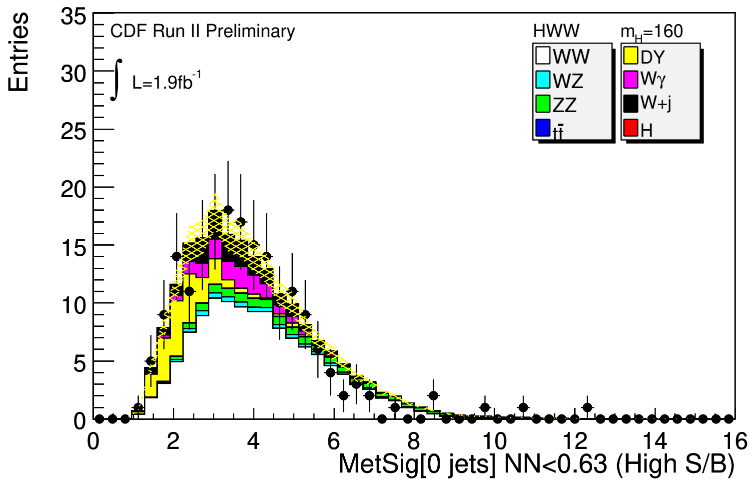


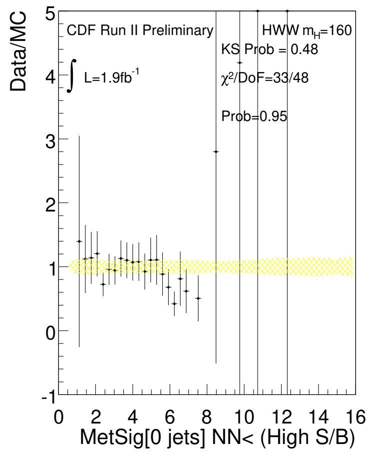


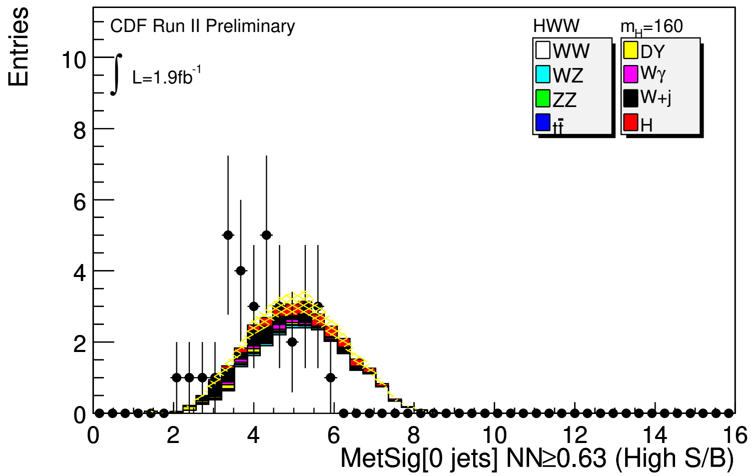


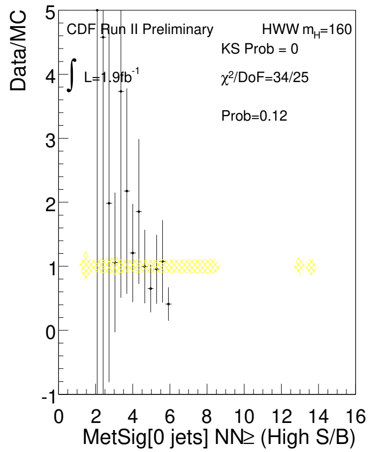


MetSig: High S/B 0 Jet









Summary

- **MetSig** The disagreement is in a region where signal is dropping and looks ok in control. Impact remains in signal only and still no way to see why this should be so.
- **dPhiLeptons** Possible problems in 1 jet, but control region does not show sigificant problems. Statistics is not overwhelming.
- **dimass** The low signal, 0 jet, low S/B of concern as it is high in signal region. Control region has good agreement where the signal is expected. Z peak disagreement does not impact signal at all.